DOCUMENT RESUME

ED 220 491 TM 820 504

AUTHOR Newman, Dorothy C.; Stallings, William M.
TITLE Teacher Competency in Classroom Testing, Measurement Preparation, and Classroom Testing Practices.
PUB DATE Mar 82
NOTE 38p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New York, NY, March 20-22, 1982).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Competence; *Educational Testing; Measurement; Teacher Education Programs; Teacher Effectiveness; *Teacher Made Tests; *Teachers; *Testing
IDENTIFIERS Measurement Competency Test (Mayo); *Teacher Competencies

ABSTRACT

An assessment instrument and a questionnaire (Appendices A and B) were developed to determine how well teachers understand classroom testing principles and to gain information on the measurement preparation and classroom practices of teachers. Two hundred ninety-four inservice teachers, grades 1 through 12, from three urban school systems in Alabama, Florida and Georgia, responded by mail to the instruments. Results indicated teachers have only a moderate understanding of classroom testing principles, even though they have completed a substantial amount of measurement training. Measurement training practices do not seem to be any more effective now than they were over 15 years ago. Such findings suggest a need for qualitative change in preservice and inservice measurement training. Questionnaire results showed that teachers prefer certain test purposes, instructional objectives, and types of items for their teacher-made tests. (Author/PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED220491

TEACHER COMPETENCY IN CLASSROOM TESTING, MEASUREMENT PREPARATION, AND CLASSROOM TESTING PRACTICES

Dorothy C. Newman
Spring Hill College
Mobile, Alabama

William M. Stallings
Georgia State University
Atlanta, Georgia

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D. C. Newman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at the Annual Meeting of the
National Council on Measurement in Education
New York, March 1982

TM 820504

Teacher Competency in Classroom Testing, Measurement Preparation, and Classroom Testing Practices

Measurement of student learning in the classroom is an especially important part of teaching. As Airasian noted,"the school and classroom are evaluative environments in which testing and inference-making are both integral and essential aspects" (p. 4). Teachers use their own tests to make a variety of instructional decisions, yet little is known about teacher competency in classroom testing. This study was conducted to establish the extent of teacher preparation in measurement; to document common classroom testing practices; to determine how well teachers understand classroom testing principles; and to investigate the relationships between classroom testing competency and measurement preparation, classroom practice, and background variables.

Over the years educators have been aware of the need for teacher competency in measurement. Much of the literature on the subject points to a lack of emphasis on measurement training (Goslin, 1967; Mayo, 1967; Reynolds & Menard, 1980; Roeder, 1973; Joyce, Yarger, Howey, Harbeck, & Kluwin, 1977). Measurement educators almost unanimously agree that this adversely affects classroom practice (Green, 1975; Stanley & Hopkins, 1972).

Little has been done, however, to assess measurement competency directly. Researchers who have studied the problem, directly or indirectly, have found teacher knowledge of measurement principles to be inadequate. According to Florida's coordinators of accountability, only 25% of Florida's teachers show evidence of adequate measurement preparation, while only 10% to 20% have the ability to answer basic questions on classroom measurement principles (Hills, 1977).

Farr and Griffin built a case in support of two hypotheses.

> The first is that in general teachers have quite limited knowledge of measurement concepts. The second is that teachers are not being taught what they need to know about measurement in order to be more effective teachers. (1973, p. 19)

Research done by Reynolds and Menard (1980) indicates that teachers who are aware of the importance of measurement principles may fail to apply them in the classroom. These researchers surveyed 187 New York secondary teachers who rated 18 test construction practices (1) according to the extent they actually followed the practice and (2) according to the extent they felt the practice should be followed. On 17 of the 18 practices, mean ratings for "should follow" were significantly higher than those for "do follow". Reynolds and Menard concluded that "teachers did not appear to see testing as an important part of teaching" (p. 10).

In 1964 Mayo (1967) conducted the most extensive study of teacher measurement competency to date, a large-scale national effort sponsored by NCME and funded by USOE. One objective was a direct assessment of preservice teacher proficiency in several measurement areas including classroom testing. Mayo developed two forms of his Measurement Competency Test (MCT) which were taken by 2,877 graduating seniors in 86 teacher preparation institutions. Results showed that "beginning teachers do not demonstrate a very high level of measurement competency" (p. 63).

Mayo recommended that measurement courses in teacher preparation programs be improved and that "an interesting and meaningful compulsory course" (p. 57) in measurement be part of the program. He advocated more practical measurement experiences for both preservice and inservice teachers to show them the need for proficiency in measurement and to increase their commitment to obtaining it.

Mayo concluded his final report with a statement which may be as relevant today as it was then.

> It seems apparent that the high levels of measurement competency desirable for the teacher to play his evaluative, as well as his instructional role have not materialized from traditional training practices. (1967, p. 64)

Since the publication of Mayo's landmark study, few research efforts have focused on classroom testing competency or on the relationship of such competency to measurement preparation and classroom testing practices. This study addressed the need for more research in these areas.

Methods

Subjects

Subjects for the study were 294 inservice teachers from three Southern school systems: Atlanta Public Schools, Atlanta, Georgia; Escambia County Public Schools, Pensacola, Florida; and Mobile County Public Schools, Mobile, Alabama. These systems are located in moderate-to-large cities, the urban setting in which most teachers work today. Of the 294 teachers, grades one through twelve, 96 were from Atlanta, 109 from Escambia County, and 89 from Mobile County. Almost equally represented were elementary (50.7%) and secondary (49.3%) teachers. There were more women (80.6%) than men (18.4%), more whites (58.2%) than nonwhites (40.5%), and a wide range of ages.

Instruments

The two research instruments, the Questionnaire for Inservice Teachers and the Teacher-Made Tests Assessment Instrument (see Appendixes A and B), were modeled after Mayo's Questionnaire for Seniors in Teacher-Preparation Programs and Measurement Competency Test (1967). Items on the questionnaire were constructed to obtain information on teacher background, preparation in educational measurement, and classroom testing practices. The Teacher-Made Tests

Assessment Instrument (TTAI) was developed to measure teacher competency in classroom testing, i.e., teacher knowledge and understanding of classroom testing principles. Measurement competencies needed by classroom teachers and items to measure them came from several sources: Mayo's Checklist of Measurement Competencies (1967); Mayo's Measurement Competency Test (1967); standard measurement texts by Green (1975), Gronlund (1973), Stanley and Hopkins (1972), TenBrink (1974), and Thorndike and Hagen (1977), as well as an item bank on educational measurement compiled by NCME.

During March and April 1980, tryout forms of the questionnaire and the TTAI were field tested with 42 inservice teachers enrolled for graduate work at the University of South Alabama. Based on these findings, final revisions of the instruments resulted in a 17-item questionnaire and a 30-item TTAI.

Data Collection and Analyses

In May 1980 the survey, consisting of the questionnaire and TTAI, was mailed to a random sample of 500 regular classroom teachers, grades one through twelve, in each of the three participating school systems. Each mailing included a cover letter along with an addressed, postage paid reply envelope. Given the number of eligible teachers in each system, the minimum sample sizes needed to insure a precision of $\pm$ 5% with 95% confidence for sampling attributes were 350 in Atlanta, 322 in Escambia County, and 341 in Mobile County. However, because response rates for mailed surveys rarely reach 100%, sample sizes were increased to 500 in each system. Such an oversampling procedure helps to increase the sampling precision (the associated issue of nonresponse bias is discussed below). Although responses were anonymous, surveys were coded for follow-up. Three weeks after the initial mailing, postcard reminders were sent to nonrespondents.

By early July 1980, 322 responses (21.5%) had been received, but only 294 questionnaires (19.6%) and 280 TTAIs (18.7%) were usable. Due to the response rates and concern over possible nonresponse bias, a telephone survey of 82 randomly selected nonrespondents in Mobile was conducted. Chi-square analyses showed no significant differences (.05 level) between respondents and nonrespondents on eight of nine demographic variables (teaching experience, school level, teaching field, highest degree, recency of highest degree, current degree enrollment, age, and sex). The two groups differed in racial composition, however, raising the possibility of nonresponse bias due to the underrepresentation of nonwhites in the sample. Therefore, study results dealing with race should be interpreted with caution if not disregarded altogether.

Data analyses, done for each system separately and for the total sample, included descriptive statistics and item analysis information for TTAI scores, as well as tables displaying questionnaire responses in percentages. Correlational and analysis of variance techniques were used to examine relationships between variables.

Results

Shown in the attached tables, study results fall into three categories: questionnaire, TTAI, and correlational results. The first set of tables summarizes responses to questionnaire items 9-12 dealing with measurement preparation variables (Tables 2-5) and responses to questionnaire items 13-17 dealing with classroom testing practices variables (Tables 6-10). Table 1 gives standard errors for percentages in those tables. The second set of tables shows descriptive statistics, reliability information, and item analysis data for the 30-item TTAI (Tables 11-16). The third set of

tables shows correlations between TTAI scores and measurement preparation variables (Table 17), classroom testing practices variables (Tables 18-20), and background variables (Table 21). Results were surprisingly consistent across school systems, a consistency supporting certain patterns and trends in the data.

Discussion

The inservice teachers in this study had taken a substantial amount of measurement training (see Table 2). More than three-fourths had completed one or more courses and only 12.9% had received no measurement training. These results provide a sharp contrast to results of earlier studies. In Mayo's study of graduating education students (1967, p. 29), a much larger percentage (33.9%) had completed measurement work as part of another course, about the same percentage (35.0%) had completed one measurement course, a much smaller percentage (only 1.0%) had completed more than one course, and almost a third (30.1%) had completed no measurement work.

These data suggest a trend toward more measurement work in teacher training programs in recent years. However, the difference in amount of measurement work completed by respondents in Mayo's study and this study may be attributed to the differences in coursework opportunities for preservice and inservice teachers. Even so, the large percentage of inservice teachers in this study (43.9%) with more than one measurement course is noteworthy. Regardless of the academic level of this work, it suggests a concern on the part of teachers and teacher educators for improving measurement training. The fact that such a small percentage (5.8%) had completed their measurement training as part of another course also seems to indicate a change in the type of measurement work emphasized for teachers since Mayo's study (1967). This apparent shift has been toward completion of

whole courses rather than completion of measurement units in other courses such as educational psychology.

A concern for upgrading the measurement competency of classroom teachers is also evidenced by an increased emphasis on required training (see Table 4). More than three-fourths of the 256 teachers with measurement training were required to take at least some of that work, a finding at variance with the lack of emphasis on required measurement work reported in earlier studies (Joyce et al., 1977; Roeder, 1973).

Unfortunately, the increase in amount of measurement training and concern for upgrading measurement competency have not translated into high performance levels for teachers on the TTAI (see Table 11). Teachers in this study obtained a correct response rate of 53.7% on the TTAI, demonstrating only moderate understanding of classroom testing principles.

High-scoring teachers demonstrated more ability with technical and statistical concepts than low-scoring teachers, although teachers as a group exhibited a better understanding of general principles than of the more technical aspects of testing. This is unfortunate, for most measurement educators would agree that an understanding of technical testing concepts helps produce more positive attitudes, enhance professional growth, and improve classroom testing practices. There is some question, however, as to the efficacy of traditional measurement coursework in accomplishing these goals.

In this study, although teachers with measurement training obtained better scores on the TTAI than teachers without training, neither group performed at a very high level. Teachers with training averaged 54.6% correct responses; teachers without training averaged only 48.0%. Mayo observed the same trend with beginning teachers and concluded that traditional measurement training had not adequately prepared teachers to be classroom evaluators

(1967, p. 64). Results of this study show that the situation has not improved dramatically since Mayo's observations. Simply requiring more of the traditional type of coursework has not solved the problem of inadequate measurement preparation. Instead, it seems that qualitative changes in measurement training are indicated. Maybe it is time to plan more teacher-oriented, classroom-based measurement courses.

Some of the most interesting findings of this study include the questionnaire results dealing with classroom testing practices, results that provide potentially useful information for planning a more relevant measurement course. These results also tend to confirm some of the widely accepted beliefs about classroom testing practices that are more often assumed than documented. It was found that teachers use teacher-made tests extensively and spend a substantial amount of professional time constructing, administering, scoring, and interpreting them (see Tables 6 and 7). The majority of teachers construct more than half of the tests used in their classes and spend more than 10% of their professional time on them. These results are similar to those of Reynolds & Menard who found that teachers construct an average of 73.68% of their test questions (1980, p. 6; 12).

Teachers definitely prefer certain test purposes, instructional objectives, and types of items for their classroom tests. The three most popular purposes, diagnosing strengths and weaknesses, assessing academic achievement, and assessing mastery of instructional units,are also major teaching tasks (see Table 8). A majority also cited three other common responsibilities that are popular reasons for using teacher-made tests: assessing growth, assigning grades, and planning instruction. Correlations showed that four of the six most popular test purposes--assessing achievement, assessing mastery, assigning grades, and planning instruction--are related to TTAI scores. Teachers

who use their classroom tests for these purposes performed better on the TTAI (see Table 19). It is not immediately evident why these four purposes are discriminating and the other two are not, but these relationships offer an interesting avenue for future research.

The popularity of the three types of instructional objectives seems merely to reflect the emphasis placed on each by educators (see Table 9). Cognitive objectives are most popular, for "the cognitive domain is that on which most instruction and measurement focus" (Green, 1975, p. 16). Assessing affective objectives probably is not as popular because "the appraisal of feelings, interests, and attitudes has been grossly neglected in education" (Stanley & Hopkins, 1972, p. 301). The lack of emphasis on measuring psychomotor outcomes is evidenced the topic's ommission from most measurement texts and courses. In fact, it is interesting that the percentages of teachers reporting the use of affective and psychomotor objectives are as high as they are.

Along with preferences for some test purposes and types of instructional objectives, teachers have a tendency to favor certain types of items for their teacher-made tests (see Table 10). It is evident that objective items are the most popular among teachers. Essay items are used by only a fifth of the teachers. The rankings of types of items from most to least used were identical across school systems: (1) completion, (2) multiple choice, (3) matching, (4) true-false, (5) short answer, (6) calculation, and (7) essay.

These rankings differed from those reported by Reynolds and Menard (1980, p. 6;13). Teachers in their sample ranked the item formats (1) short answer, (2) multiple choice, (3) essay, (4) matching, and (5) true-false. Differences in rankings may be due to differences in the populations studied (this study included elementary teachers). These discrepancies provide another area for future research.

Correlations showed that classroom testing competency as measured by the TTAI is related to two classroom testing practices variables: number of test purposes and number of types of items (see Table 18). Teachers who use teacher-made tests for a greater number of purposes and who write more types of items for their tests performed better on the TTAI. At the same time, no relationship is evident between classroom testing competency and either the percentage of tests made by teachers or the time spent on them. Therefore, it seems that variety in classroom testing practices (more purposes and items) is a priority with high-scoring teachers regardless of the effort and time devoted to testing activities.

Several teacher background variables were correlated with classroom testing competency. Younger teachers, teachers with higher degrees , and teachers with less teaching experience tended to score better on the TTAI (see Table 21). Since recency of measurement training is also related to classroom testing competency (see Table 17), recency is probably at least a partial explanation for the competency manifested by teachers who are younger, less experienced, and better educated.

Educational Importance

According to the results of this study, there has been little change in the status of teacher measurement competency since Mayo's landmark study of beginning teachers (1967). The current level of teacher understanding of classroom testing principles is not high. Teachers manifested only a moderate degree of competency in classroom testing as measured by the TTAI. These results call into question the adequacy of teacher preparation for classroom testing responsibilities. Measurement training practices do not seem to be any more effective now than they were over a decade and a half ago (Mayo, 1967). Even so, teachers use their tests extensively in the classroom.

It is obvious that a need still exists to improve preservice and inservice measurement training. One possibility to meet the need is for state departments of education to require measurement training for teacher certification and to include measurement competencies on required teacher tests. Another possibility is for teacher preparation institutions and/or school systems to require measurement coursework for preservice and inservice teachers. However, rather than merely requiring more of the same type of traditional measurement training, which has not brought a desired level of teacher competency, a change in the qualitative nature of measurement work seems indicated.

Mayo (1967) advocated increasing the relevance of measurement training for teachers, a recommendation that is probably more appropriate today than at any previous time. The results of this study should be useful to measurement educators in planning more relevant and effective coursework for improving classroom practice. Finding ways to make coursework more meaningful to teachers is a promising avenue for future research.

References

Airasian, P. W. A perspective on the uses and misuses of standardized achievement tests. Measurement in Education, 1979, 10(3), 1 - 12.

Farr, R., & Griffin, M. Measurement gaps in teacher education. Journal of Research and Development in Education, 1973, 7(1), 19 - 28.

Goslin, D. A. Teachers and testing. New York: Russell Sage Foundation, 1967.

Green, J. A. Teacher-made tests (2nd ed.). New York: Harper & Row, 1975.

Gronlund, N. E. Preparing criterion-referenced tests for classroom instruction. New York: Macmillan, 1973.

Hills, J. R. Coordinators of accountability view teachers' measurement competence. Florida Journal of Educational Research, 1977, 19, 39 - 44.

Joyce, B. R., Yarger, S. J., Howey, K. R., Harbeck, K. M., & Kluwin, T. N. Preservice teacher education. Washington, D. C.: Office of Education, 1977. (ERIC Document Reproduction Service No. ED 146 120)

Mayo, S. T. Preservice preparation of teachers in educational measurement (Report No. 5-0807). Chicago: Loyola University, December 1967.

Newman, D. C. Teacher competency in classroom testing, measurement preparation, and classroom testing practices (Doctoral dissertation, Georgia State University, 1981).

Reynolds, W. M., & Menard, K. A. An examination of teachers' test construction practices. Paper presented at the annual meeting of the National Council on Measurement in Education, Boston, April 1980.

Roeder, H. H. Teacher education curricula--Your final grade is F. Journal of Educational Measurement, 1973, 10(2), 141 -143.

Stanley, J. C., & Hopkins, K. D. Educational and psychological measurement and evaluation. Englewood Cliffs, N. J.: Prentice-Hall, 1972.

TenBrink, T. D. Evaluation: A practical guide for teachers. New York: McGraw-Kill, 1974.

Thorndike, R. L., & Hagen, E. P. Measurement and evaluation in psychology and education (4th ed.). New York: Wiley, 1977.

Appendixes

A. Questionnaire for Inservice Teachers

B. Teacher-Made Tests Assessment Instrument (TTAI)

APPENDIX A

QUESTIONNAIRE FOR INSERVICE TEACHERS

Directions: Please PRINT the requested information or CHECK the appropriate option(s). Your answers will remain ANONYMOUS. THANK YOU for your participation.

1. Number of years of teaching experience including this year_____

2. Grade(s) currently teaching (If you check more than one, circle the grade of major responsibility.)

___Grade 1 ___Grade 3 ___Grade 5 ___Grade 7 ___Grade 9 ___Grade 11

___Grade 2 ___Grade 4 ___Grade 6 ___Grade 8 ___Grade 10 ___Grade 12

___Other (Specify)_____

3. Teaching field(s) (If you check more than one, circle the field of major responsibility.)

___General Early Childhood	___Social Science	___Physical Education
___General Middle Childhood	___Mathematics	___Vocational/Career Educ.
___General Elementary	___Science	___Art
___Reading	___Foreign Language	___Music
___Language Arts	___Home Economics	___Other (Specify)
___English	___Health Education	_____

4.

Degree(s) Held	Major	Year Awarded
_____	_____	_____
_____	_____	_____

5. Are you currently enrolled in a degree program? ___YES ___NO

If YES, Degree? _____ Major? _____

6. Age_____

7. Sex_____

8. To which ethnic or racial group do you belong? (Check one)

___American Indian

___Asian American

___Black American

___Hispanic American

___White American

___Other (Specify)_____

9. How much coursework in testing and measurement have you completed? (Check as many as apply)

___None

___Part of another course (Specify course name)_____

___One college course

___More than one college course

___An inservice course (one offered by a school system)

If you have completed NO COURSEWORK in testing and measurement, GO to question 13. SKIP questions 10-12.

10. On what level did you take your measurement coursework? (Check as many as apply)

___Undergraduate

___Graduate

___Inservice (course offered by a school system)

___Other (Specify)_____

11. Was your measurement coursework required?_____

12. In what year did you complete your most recent measurement coursework?_____

13. What percentage of the tests you use in your classroom do you construct yourself? (Check one)

____ 0 - 25%

____26 - 50%

____51 - 75%

____76 -100%

14. What percentage of your professional time (both in and out of school) do you spend constructing, administering, scoring, and interpreting your teacher-made tests? (Check one)

___Less than 5%

____ 5 - 10%

____11 - 20%

___More than 20%

15. For what purposes do you use your teacher-made tests? (Check as many as apply)

____Diagnosing strengths and weaknesses

____Assessing academic achievement

____Assessing mastery/instructional units

____Assessing academic progress/growth

____Motivating learning

___Assigning grades

____Determining Group Membership

___Planning instruction

____Choosing curriculum

____Other (Specify) _____

16. What types of objectives do you test on your teacher-made tests? (Check as many as apply)

____Cognitive objectives ____Affective objectives ____Psychomotor objectives

17. What types of items do you write for your teacher-made tests? (Check as many as apply)

__True/False

__Matching

__Multiple Choice

___Completion (fill-in-the-blank)

___Short Answer(one/two paragraphs)

___Essay (one/two pages)

___Calculation Problems

___Other (Specify)

_____

APPENDIX B

TEACHER-MADE TESTS ASSESSMENT INSTRUMENT

In the blank beside each question, PRINT the letter of the answer you believe to be correct. Thank you.

___1. What is the main advantage of using teacher-made tests rather than standardized tests?

A. They are more reliable for evaluating individual differences.

B. They provide a better measure of the student's understanding of important principles.

C. They provide more valid measures of the teacher's specific objectives.

D. They are simpler to administer and score.

___2. The most useful statements of instructional objectives are those which

A. relate skills and knowledge to behavioral outcomes.

B. can be used as the basis for objective tests.

C. are specific in terms of the knowledge they involve.

D. are equally appropriate for all students in a class.

___3. The most important reason to develop a test plan is to insure the test's

A. predictive validity.

B. content validity.

C. test-retest reliability.

D. internal consistency.

___4. In constructing an achievement test, which of the following should come first?

A. Writing the test items.

B. Composing the instructions.

C. Examining tests constructed by other teachers.

D. Defining instructional objectives.

___5. An objective test is one in which

A. teacher judgment in test construction is eliminated.

B. subjectivity in scoring is eliminated.

C. items relate directly to course objectives.

D. there is valid coverage of course objectives.

___6. Which objective test items are most likely to measure ability to analyze and synthesize information?

A. Matching items.

B. Multiple choice items.

C. Fill-in-the-blank items.

D. Recognition items.

___7. Test items known as <u>matching</u> items should be used for measuring the ability to

A. understand principles.

B. analyze complex relationships.

C. identify facts.

D. synthesize information.

___8. Why is this a <u>poor</u> true-false item? "New York is the most important city in the United States."

A. It is too easy.

B. It is too brief.

C. It is too factual.

D. It is ambiguous.

___9. In evaluating a history essay test which of the following is <u>not</u> a desirable practice?

A. Deducting points for poor spelling and handwriting.

B. Preparing a scoring key and standards in advance.

C. Removing students' names from the papers.

D. Scoring one question on all papers before going to the next question.

___10. To increase the reliability of evaluating an essay test the teacher should

A. grade the papers on the normal curve.

B. have several optional essay questions on the test.

C. assign weights to each essential part of an answer.

D. divide up the papers among different teachers for grading.

___11. Essays are more appropriate than objective test items to measure

A. organization of ideas.

B. knowledge of basic facts.

C. application of knowledge.

D. comprehension of underlying principles.

___12. What is the main advantage of using an objective test rather than an essay test?

A. It is easier to administer.

B. It is quicker to construct.

C. It provides for comparisons among individual scores.

D. It provides for a more complete sampling of content.

___13. Which educational outcome is classified in the affective domain?

A. Improving reading comprehension.

B. Developing handwriting skills.

C. Developing an interest in music.

D. Understanding a classification of interests.

___14. The type of validity that should most concern a teacher constructing an achievement test is

A. predictive validity.

B. construct validity.

C. content validity.

D. concurrent validity.

___15. A correlation of .95 between two sets of scores for the same test given to the same class indicates that the scores are

A. highly reliable.

B. highly valid.

C. not very stable.

D. not very differentiating.

___16. The statement, "This test is valid," will be most improved by specifying

A. the purpose for which the test is valid.

B. the procedures for scoring.

C. the conditions for administration.

D. a definition of the word "valid".

___17. Because no test is perfectly reliable, a student's score should be regarded as

A. having little meaning unless very high or very low.

B. indicating a point in the range near which the student's true score probably falls.

C. indicating only that the student has more or less ability than the average student.

D. having little validity in estimating the student's achievement.

___18. On a criterion referenced test, each student's score is compared to

A. a standard of performance.

B. the mean score of the test.

C. the true score of the student.

D. the performance of the class as a whole.

___19. Which of the following is <u>not</u> a characteristic of criterion referenced testing?

A. A well-specified domain of learning tasks.

B. Instructional objectives clearly defined in behavioral terms.

C. Test items selected on the basis of how well they reflect instructional objectives.

D. Test performance reported in terms of relative position in a group.

___20. A criterion referenced test is more appropriate than a norm referenced test for decisions about

A. admission to college.

B. determination of mastery.

C. prediction of college success.

D. placement in a gifted program.

___21. In which case is a teacher most justified in setting the criterion score at 80% for all students?

A. The class is composed of exceptional students in a program for gifted children.

B. The class has ample time to study the material and prepare for the test.

C. The questions measure knowledge and skills needed for success in the next unit.

D. The questions are essay rather than objective questions.

___22. Don made a score of 70 on his economics criterion referenced mastery test. Which information would be most helpful in describing Don's achievement in economics?

A. The test consisted of 90 questions.

B. Half of the class failed the test.

C. The mean score for the class was 80.

D. The minimum passing score on the test was 75.

___23. In general, adding items to a test will make the test more

A. reliable.

B. diagnostic.

C. objectives.

D. valid.

___24. Knowing the percent of students getting each item correct (item difficulty level) helps a teacher to

A. evaluate attainment of instructional objectives.

B. arrange items in order of difficulty on a test.

C. revise a series of test items.

D. accomplish all of the above.

___25. On one test item the correct answer was marked by 40% of students in the top fourth of the class and by 65% of students in the lowest fourth of the class. The item's discriminating power is

A. almost perfect.

B. negative.

C. positive.

D. stable.

___26. Which correlation is the strongest?

A. r = -.82

B. r = +.72

C. r = -.54

D. r = +.33

Listed below are the options for questions 27 through 30. Options <u>can</u> be used more than once.

A. the mean.

B. the median.

C. the standard deviation.

D. the standard error of measurement.

___27. An index indicating a range of scores within which the student's true score most probably falls is . . .

___28. The point on a measurement scale above which and below which there are 50% of the cases is . . .

___29. An average that is useful when a distribution of scores includes a few very high or low scores is . . .

___30. An index indicating the amount of variability in a distribution of test scores is . . .

During the summer of 1980, after the project is completed, would you like for us to send you:

(1) a copy of the study results? ____YES ____NO

(2) a bookmark as a small token of our appreciation for your participation? ____YES ____NO

If you answered YES to either question, please fill in your mailing address. It is not necessary to give your name. This information is confidential.

Mailing Address _____

_____

- - - -

TTAI Key

1. C	11. A	21. C
2. A	12. D	22. D
3. B	13. C	23. A
4. D	14. C	24. D
5. B	15. A	25. B
6. B	16. A	26. A
7. C	17. B	27. D
8. D	18. A	28. B
9. A	19. D	29. B
10. C	20. B	30. C

Table 1

Standard Errors for Selected Percentages and Sample Sizes

	Percentages										
	99	95	90	85	80	75	70	65	60	55	50
n	1	5	10	15	20	25	30	35	40	45	50
75	1.2	2.5	3.5	4.2	4.6	5.0	5.3	5.5	5.7	5.8	5.8
100	1.0	2.2	3.0	3.6	4.0	4.4	4.6	4.8	4.9	5.0	5.0
150	.8	1.8	2.5	2.9	3.3	3.5	3.8	3.9	4.0	4.1	4.1
200	.7	1.5	2.1	2.5	2.8	3.1	3.2	3.4	3.5	3.5	3.5
250	.6	1.4	1.9	2.3	2.5	2.7	2.9	3.0	3.1	3.2	3.2
300	.6	1.3	1.7	2.1	2.3	2.5	2.7	2.8	2.8	2.9	2.9

Table 2

Amount of Measurement Coursework

Amount	Atlanta (n = 96) %	Escambia Co. (n = 109) %	Mobile Co. (n = 89) %	Total (n = 294) %
Part of Another Course	4.2	9.2	3.4	5.8
One Course	31.3	36.7	31.5	33.3
More Than One Course	49.0	38.5	44.9	43.9
Inservice Only	4.2	4.6	3.4	4.1
Subtotals	88.5	89.0	83.1	87.1
None	11.5	11.0	16.9	12.9

Table 3

Level of Measurement Coursework

Level	Atlanta (n = 85) %	Escambia Co. (n = 97) %	Mobile Co. (n = 74) %	Total (n = 256) %
Undergraduate	12.9	43.3	40.5	32.4
Graduate	44.7	27.8	16.2	30.1
Undergraduate and Graduate	28.2	15.5	36.5	25.8
Inservice Only	4.7	5.2	4.1	4.7
No Answer	9.4	8.2	2.7	7.0

Note. Table is based on data from 256 teachers with measurement training.

Table 4

Requirement Status of Measurement Work

Status	Atlanta (n = 85) %	Escambia Co. (n = 97) %	Mobile Co. (n = 74) %	Total (n = 256) %
All or Part Required	76.5	73.7	86.5	77.7
None Required	11.8	20.6	9.5	14.5
No Answer	11.8	7.2	4.1	7.8

Note. Table is based on data from 256 teachers with measurement training.

Table 5

Recency of Measurement Coursework

Years	Atlanta (n = 85) %	Escambia Co. (n = 97) %	Mobile Co. (n = 74) %	Total (n = 256) %
1975 - 80	34.1	35.1	45.9	37.9
1970 - 74	30.6	21.6	27.0	26.2
1965 - 69	8.2	11.3	6.8	9.0
1960 - 64	4.7	5.2	1.4	3.9
Pre - 1960	7.1	10.3	12.2	9.8
No Answer	15.3	16.5	6.8	13.3

Note. Table is based on data from 256 teachers with measurement training.

Table 6

Percentage of Classroom Tests Made by Teachers

Percentages	Atlanta (n = 96) %	Escambia Co. (n = 109) %	Mobile Co. (n = 89) %	Total (n = 294) %
0 - 25%	20.8	12.8	18.0	17.0
26 - 50%	33.3	16.5	16.9	22.1
51 - 75%	18.8	19.3	22.5	20.1
76 -100%	25.0	46.8	42.7	38.4
No Answer	2.1	4.6	.0	2.4

Table 7

Percentage of Time Spent on Teacher-Made Tests

Percentages	Atlanta (n = 96) %	Escambia Co. (n = 109) %	Mobile Co. (n = 89) %	Total (n = 294) %
Less Than 5%	7.3	13.8	9.0	10.2
5 - 10%	22.9	18.3	21.3	20.7
11 - 20%	31.3	34.9	30.3	32.3
Greater Than 20%	36.5	28.4	37.1	33.7
No Answer	2.1	4.6	2.2	3.1

Table 8

Purposes of Teacher-Made Tests

Purpose	Atlanta (n = 96) %	Escambia Co. (n = 109) %	Mobile Co. (n = 89) %	Total (n = 294) %
Diagnosis	84.4	78.0	82.0	81.3
Assessing Achievement	64.6	77.1	82.0	74.5
Assessing Mastery	70.8	76.1	76.4	74.5
Assessing Growth	60.4	61.5	75.3	65.3
Assigning Grades	33.3	73.4	78.7	61.9
Planning Instruction	70.8	51.4	56.2	59.2
Motivating Learning	45.8	49.5	40.4	45.6
Determining Groups	47.9	24.8	25.8	32.7
Choosing Curriculum	19.8	16.5	10.1	15.6
No Answer	3.1	2.8	.0	2.0

Note. Categories are not exclusive.

Table 9

Types of Objectives for Teacher-Made Tests

Type	Atlanta (n = 96) %	Escambia Co. (n = 109) %	Mobile Co. (n = 89) %	Total (n = 294) %
Cognitive	84.4	88.1	92.1	88.1
Affective	54.2	58.7	60.7	57.8
Psychomotor	15.6	26.6	24.7	22.5
No Answer	10.4	6.4	4.5	7.1

Note. Categories are not exclusive.

Table 10

Types of Items Written for Teacher-Made Tests

Type	Atlanta (n = 96) %	Escambia Co. (n = 109) %	Mobile Co. (n = 89) %	Total (n = 294) %
Completion	86.5	76.1	84.3	82.0
Multiple Choice	75.0	74.3	79.8	76.2
Matching	72.9	74.3	76.4	74.5
True - False	55.2	62.4	66.3	61.2
Short Answer	49.0	58.7	51.7	53.4
Calculation	47.9	40.4	43.8	43.9
Essay	19.8	19.3	24.7	21.1
No Answer	5.2	4.6	1.1	3.7

Note. Categories are not exclusive.

Table 11

Descriptive Statistics and Sampling Precision for TTAI Results

Sample	n	$\bar{X}$	s	SE	95% C.I. for $\bar{X}$
Atlanta	89	14.6	4.2	.44	13.7 to 15.5
Escambia Co.	102	16.8	4.9	.49	15.8 to 17.8
Mobile Co.	89	16.7	4.7	.50	15.7 to 17.7
Total Sample	280	16.1	4.7	.28	15.5 to 16.6

Table 12

TTAI Reliability

Index	Atlanta	Escambia Co.	Mobile Co.	Total
KR-20	.66	.77	.74	.74

Table 13

TTAI Standard Errors of Measurement

Index	Atlanta	Escambia Co.	Mobile Co.	Total
SEM	2.43	2.36	2.39	2.40

Table 14

Item Analysis Data for TTAI: Total Sample

Item	Diff.[a]	r_{pb}[b]	Item	Diff.[a]	r_{pb}[b]
1	80	30	16	62	33
2	58	37	17	81	35
3	60	13	18	60	41
4	88	32	19	55	42
5	44	47	20	76	34
6	47	32	21	40	42
7	65	16	22	37	40
8	89	24	23	25	39
9	44	36	24	58	37
10	70	32	25	58	46
11	40	32	26	18	28
12	54	24	27	24	47
13	36	54	28	55	37
14	72	19	29	17	09
15	47	37	30	46	42

Note. Decimal points are omitted in table.

[a]Difficulty Index.

[b]Point Biserial Correlation.

Table 15

Difficulty Indexes for Easy and Hard TTAI Items

Item	Atlanta	Escambia Co.	Mobile Co.	Total
1	72	84	84	80
4	88	90	84	88
7	--	71	--	--
8	88	93	85	89
10	--	78	74	70
12	--	--	71	--
13	25	--	--	--
14	71	74	72	72
16	--	70	--	--
17	79	85	80	81
18	--	--	73	--
20	80	--	89	76
23	26	28	21	25
26	17	16	21	18
27	16	30	25	24
29	20	13	18	17

Note. Decimal points are omitted in table.

Table 16

Point Biserial Correlations for Discriminating Items on the TTAI

Item	Atlanta	Escambia Co.	Mobile Co.	Total
2	--	--	46	--
4	--	--	48	--
5	--	48	54	47
8	--	--	43	--
9	--	--	50	--
11	--	--	46	--
13	51	55	50	54
14	42	--	--	--
15	--	42	--	--
17	--	40	--	--
18	--	49	--	41
19	--	50	43	42
20	--	42	--	--
21	40	54	--	42
22	--	46	--	40
23	48	--	46	--
25	--	52	41	46
26	40	--	--	--
27	--	50	50	47
28	41	45	--	--
30	41	42	41	42

Note. Decimal points are omitted in table.

Table 17

Correlations Between TTAI Scores and Measurement Preparation Variables

Variables	Atlanta	Escambia Co.	Mobile Co.	Total
	TTAI Scores			
Amount of Measurement Work	.13 (n = 86)	.04 (n = 97)	.20* (n = 86)	.10 (n = 269)
Recency of Measurement Work	.27* (n = 68)	.18 (n = 97)	.07 (n = 69)	.15* (n = 234)

*p < .05.

Table 18

Correlations Between TTAI Scores and Classroom Testing Practices Variables

Variables	Atlanta	Escambia Co.	Mobile Co.	Total
	TTAI Scores			
Percentage of Tests Made by Teachers	.01 (n = 87)	.10 (n = 98)	.04 (n = 87)	.09 (n = 274)
Percentage of Time Spent on Teacher-Made Tests	.08 (n = 87)	.02 (n = 97)	.10 (n = 87)	.06 (n = 271)
Number of Test Purposes for Teacher-Made Tests	.09 (n = 86)	.27** (n = 99)	.28** (n = 89)	.22** (n =274)
Use of Various Types of Objectives	.26* (n = 75)	.08 (n = 90)	-.12 (n = 82)	.08 (n = 247)
Number of Item Types on Teacher-Made Tests	.01 (n = 84)	.22* (n = 98)	.14 (n = 88)	.13* (n = 270)

*p < .05.

**p < .01.

Table 19

Correlations Between TTAI Scores and Use of Test Purposes

Test Purposes	Atlanta (n = 86)	Escambia Co. (n = 99)	Mobile Co. (n = 89)	Total (n = 274)
	TTAI Scores			
Diagnosis	.04	.05	.14	.07
Assessing Achievement	.30**	.19	.17	.24**
Assessing Mastery	-.01	.26**	.28**	.19**
Assessing Growth	-.01	.17	.10	.11
Assigning Grades	.01	.12	.21*	.19**
Planning Instruction	.17	.22*	.21*	.16**
Motivating Learning	-.09	.10	-.11	-.02
Determining Groups	.09	.09	.09	.04
Choosing Curriculum	-.09	.10	.21*	.04

*$p < .05$.

**$p < .01$.

Table 20

Correlations Between TTAI Scores and Use of Types of Items

Types	Atlanta (n = 84)	Escambia Co. (n = 98)	Mobile Co. (n = 89)	Total (n = 271)
	TTAI Scores			
Completion	.07	.07	.03	.03
Multiple Choice	-.08	.11	.08	.04
Matching	.04	.16	-.04	.06
True-False	-.14	.03	-.03	-.02
Short Answer	.06	.04	.14	.08
Calculation	.13	.33**	.30**	.24**
Essay	-.01	-.04	-.02	-.02

**$p < .01$.

Table 21

Correlations Between TTAI Scores and Teacher Background Variables

Variables	Atlanta	Escambia Co.	Mobile Co.	Total
	TTAI Scores			
Years of Teaching Experience	.10 (n = 83)	-.13 (n = 101)	-.29** (n = 89)	-.14* (n = 273)
Highest Degree	.16 (n = 87)	.17 (n = 100)	.24* (n = 88)	.13* (n = 275)
Recency of Highest Degree	.02 (n = 86)	.18 (n = 97)	.12 (n = 84)	.09 (n = 267)
Current Degree Enrollment	-.20 (n = 89)	.13 (n = 102)	.24* (n = 89)	.09 (n = 280)
Age	.08 (n = 82)	-.20* (n = 99)	-.29** (n = 86)	-.15** (n = 267)
Sex	-.17 (n = 88)	-.03 (n = 101)	.07 (n = 88)	.00 (n = 277)
Race	.21* (n = 86)	.15 (n = 101)	.19 (n = 87)	.26** (n = 274)

*$p < .05$.

**$p < .01$.